Remote Sensing of Environment 265 (2021) 112636

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters

Zhuo Zheng [a], Yanfei Zhong [a,b,*], Junjue Wang [a], Ailong Ma [a], Liangpei Zhang [a,b]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430074, China
[b] Hubei Provincial Engineering Research Center of Natural Resources Remote Sensing Monitoring, Wuhan University, Wuhan 430079, China

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Building damage assessment
Change detection
Disaster response
Deep learning
Remote sensing
OBIA

ABSTRACT

Sudden-onset natural and man-made disasters represent a threat to the safety of human life and property. Rapid and accurate building damage assessment using bitemporal high spatial resolution (HSR) remote sensing images can quickly and safely provide us with spatial distribution information and statistics of the damage degree to assist with humanitarian assistance and disaster response. For building damage assessment, strong feature representation and semantic consistency are the keys to obtaining a high accuracy. However, the conventional object-based image analysis (OBIA) framework using a patch-based convolutional neural network (CNN) can guarantee semantic consistency, but with weak feature representation, while the Siamese fully convolutional network approach has strong feature representation capabilities but is semantically inconsistent. In this paper, we propose a deep object-based semantic change detection framework, called ChangeOS, for building damage assessment. To seamlessly integrate OBIA and deep learning, we adopt a deep object localization network to generate accurate building objects, in place of the superpixel segmentation commonly used in the conventional OBIA framework. Furthermore, the deep object localization network and deep damage classification network are integrated into a unified semantic change detection network for end-to-end building damage assessment. This also provides deep object features that can supply an object prior to the deep damage classification network for more consistent semantic feature representation. Object-based post-processing is adopted to further guarantee the semantic consistency of each object. The experimental results obtained on a global scale dataset including 19 natural disaster events and two local scale datasets including the Beirut port explosion event and the Bata military barracks explosion event show that ChangeOS is superior to the currently published methods in speed and accuracy, and has a superior generalization ability for man-made disasters.

1. Introduction

Rapid and accurate building damage assessment is critical for humanitarian assistance and disaster response when sudden-onset disasters happen (Gupta et al., 2019b). However, assessing the building damage can be dangerous, difficult, and slow, because of the limited communication and transportation infrastructure. Remote sensing technology is a safe and efficient way to achieve building damage assessment. High spatial resolution (HSR) remote sensing images can accurately reflect the surface of the Earth, and can rapidly provide large area observations to support building damage assessment. When based on co-registered bitemporal HSR remote sensing images, building damage assessment can be seen as a combination of two fundamental sub-tasks: building localization and damage classification. Building localization, which is also termed building extraction, has been widely studied in the remote sensing field (Liu et al., 2020), the goal of which is to assign a unique semantic label to each pixel on the pre-disaster image to indicate the building area. Then based on the building localization results on the pre-disaster image, the damage classification involves assigning a unique damage level label reflecting the degree of damage to each building instance on the post-disaster image.

Over the past few years, many researchers have tried to use moderate-resolution remote sensing images to assess building damage based on pixel-based and object-based land cover classification (Yusuf

* Corresponding author at: State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430074, China.
E-mail addresses: zhengzhuo@whu.edu.cn (Z. Zheng), zhongyanfei@whu.edu.cn (Y. Zhong).

https://doi.org/10.1016/j.rse.2021.112636
Received 16 May 2021; Received in revised form 5 July 2021; Accepted 2 August 2021
Available online 24 August 2021

et al., 2001; Yamazaki and Matsuoka, 2007). However, limited by the spatial resolution of the sensors, these methods can only assess the coarse building damage area. With the increases of the spatial resolution of the sensors, HSR remote sensing images can now be obtained to finely characterize the spatial details of each building instance, which makes it possible to assess instance-level building damage. As a result, building damage assessment based on HSR remote sensing images has gradually become the main stream approach. For example, Brunner et al. (2010) jointly used HSR optical and SAR images to extract building instances and assess instance-wise building damage via geometric parameter estimation for an earthquake disaster; and Tong et al. (2012) used IKONOS images to detect each collapsed building based on the 3D geometric changes for an earthquake disaster. These methods are mainly designed to handle the scenario of a single disaster, and they are based on the use of specific hand-crafted features for SAR and optical images in the building damage assessment (Dong and Shan, 2013). The building damage assessment problem is then modeled as a pre- and post-event change detection problem (Plank, 2014). However, the occurrence and the type of sudden-onset disasters are always unpredictable, and the design of hand-crafted features is time-consuming, which makes it difficult to quickly assess building damage over a wide scale for rapid disaster response.

Deep learning techniques, especially deep convolutional neural networks (CNNs), have achieved significant improvements in the computer vision field, and have been successfully applied to various remote sensing applications, such as object detection (Cheng et al., 2016; Zhong et al., 2018; Zheng et al., 2020a), object segmentation (Zheng et al., 2020c), land-use classification (Huang et al., 2018; Zhang et al., 2018, 2019, 2020), hyperspectral image classification (Zheng et al., 2020b), and multi-modality all-weather mapping (Zheng et al., 2021). CNNs are hierarchical feature representation learning methods, which allows us to extract high-level features from raw image data using the data-driven paradigm and apply these features to downstream tasks in an end-to-end fashion. These characteristics can significantly accelerate the whole pipeline of building damage assessment after a sudden onset disaster (Ge et al., 2020; Koshimura et al., 2020).

The CNN-based building damage assessment methods can be broadly divided into cascade network based methods and Siamese network based methods, from the perspective of the network architecture. The cascade network based approaches (Gupta et al., 2019b) use a fully convolutional network (FCN) for the building localization and a patch-based CNN for the damage classification (Valentijn et al., 2020; Lee et al., 2020). The whole algorithm is a two-step pipeline, where the first step involves predicting the pixel-wise building positions on the pre-disaster image using a building localization model. Then, in the second step, a damage classification model is applied to perform patch-wise building damage classification on the post-disaster image. For example, the xView2 baseline (Gupta et al., 2019b) implements building localization with a modified UNet architecture (Ronneberger et al., 2015) and damage classification with a two-branch ResNet-50 (He et al., 2016). However, the cascade network-based methods suffer from the knowledge gap problem between the building localization model and the damage classification model. These two models both need to recognize the buildings, but they are unable to share common knowledge with each other. This is because, in the cascade network based methods, building localization is modeled as a pixel-level classification task, whereas the damage classification is modeled as an image-level classification task. These two heterogeneous tasks respectively require a patch-based CNN and an FCN. To solve this problem, the Siamese network-based methods use the same network architecture to complete these two tasks by modeling them both as a pixel-level classification task using two FCNs. For example, Siamese-UNet (Durnov, 2020), as an xView2 Data Challenge (Gupta et al., 2019a) 1st place solution, uses two identical UNet architectures with shared weights to extract dense feature maps from the last decoder block for the pre- and post-disaster images. These two dense feature maps are then concatenated for pixel-level damage classification. Siamese-UNet adopts a two-stage training strategy, where it is first trained on pre-disaster images for the building localization, and is then trained on both pre- and post-disaster images for the damage classification based on the weights obtained in the first stage.

Although the Siamese network based methods overcome the knowledge gap problem by sharing weights, they do introduce the semantically inconsistent damage classification problem, because of the pixel-level modeling, as shown in Fig. 1(d). In principle, each building instance has only one status, as shown in Fig. 1(c). However, many weaker disasters result in buildings being partially damaged, and the damaged areas are often much fewer than the non-damaged areas. The non-damaged samples still dominate the learning procedure, which results in the pixel-level model only being able to accurately recognize the damaged area rather than an entire damaged building instance. Object-based image analysis (OBIA) integrated with patch-based CNN is commonly used to overcome this semantic inconsistency, the effectiveness of which has been widely confirmed (Blaschke, 2010; Zhang et al., 2018; Liu et al., 2021). The patch-based CNNs integrated OBIA (Zhang et al., 2018, 2019, 2020; Liu et al., 2021) mainly adopt superpixel segmentation to generate objects, but the objects are non-semantic and of an irregular geometrical shape. However, in building damage assessment, semantic inconsistency occurs in a building object that is semantic and of a regular geometrical shape, which causes the problem of it being impossible to apply these conventional OBIA methods in building damage assessment. The root of the problem lies in the fact that the current OBIA methods are simply combined with deep learning at a procedure level, without feature-level interaction.

In this paper, we propose a deep object-based semantic change detection framework, called ChangeOS, for building damage assessment

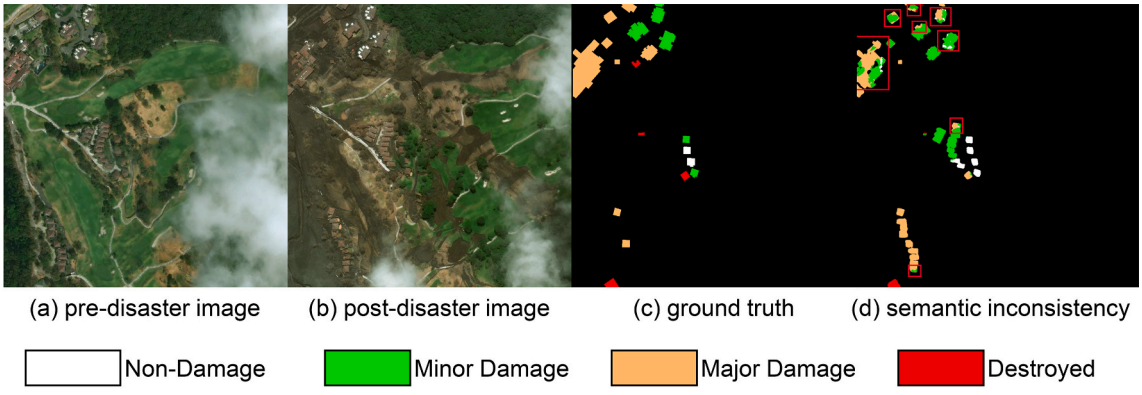

Fig. 1. The semantically inconsistent damage classification problem. The red boxes indicate the partially damaged region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

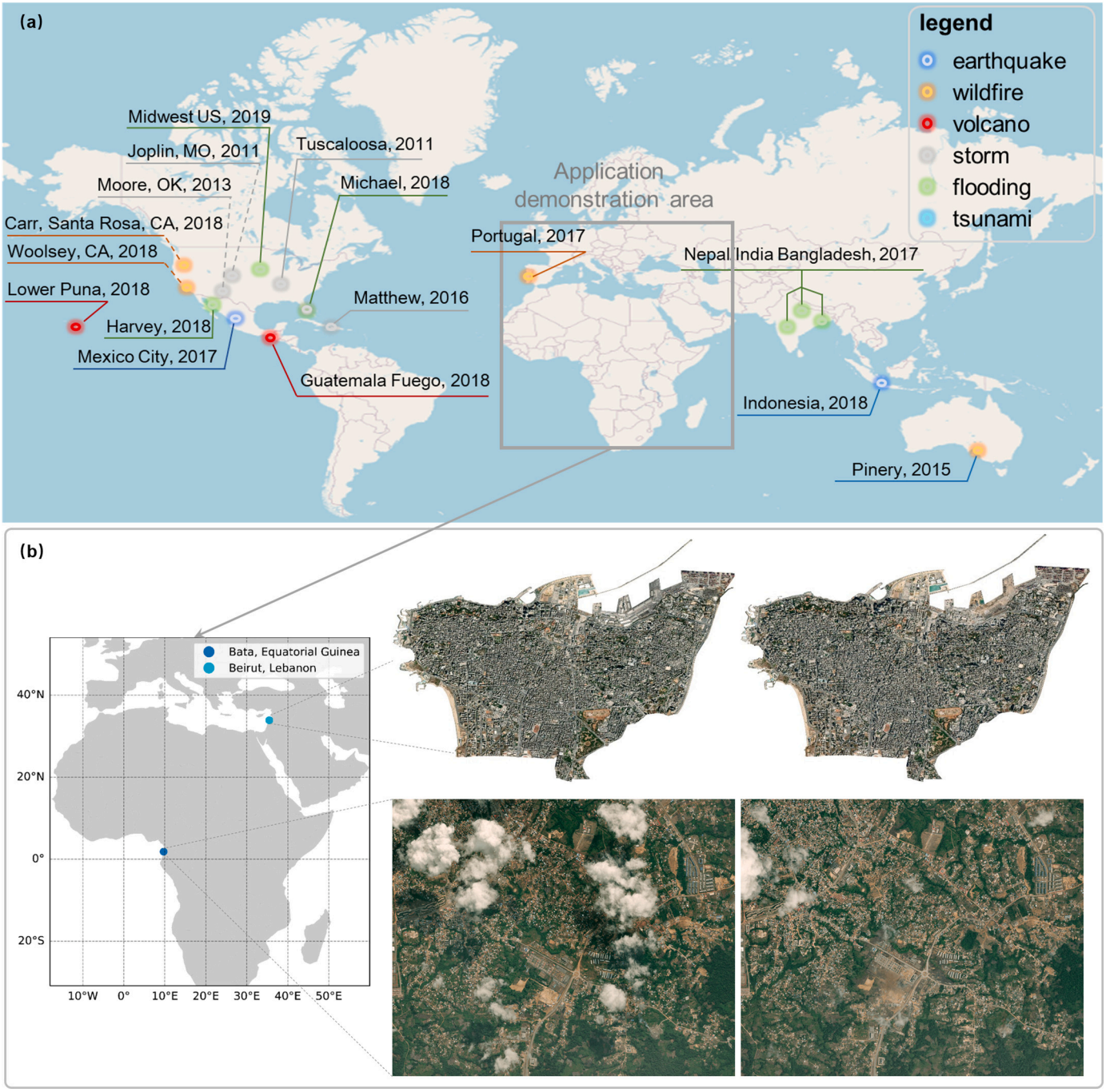

Fig. 2. The global scale study area for building damage assessment in sudden-onset disaster. (a) The distribution of the disaster events at a global scale. (b) The two extra study sites for the out-of-the-distribution explosion events.

in a disaster context using HSR bitemporal satellite images. To ensure that the OBIA is seamlessly integrate with the deep learning, we adopt an FCN to generate more accurate building objects, in place of super-pixel segmentation, and this FCN is also responsible for the pixel-level building localization. Furthermore, ChangeOS integrates building localization and damage classification into a unified framework by the use of a partial Siamese FCN architecture, thus achieving feature representation level interaction. ChangeOS has the key advantage that it can achieve end-to-end training and inference. This makes the building damage assessment model easy to tune and deploy, which significantly accelerates the whole pipeline of disaster response. The code is available at https://github.com/Z-Zheng/ChangeOS.

The rest of this paper is organized as follows. Section 2 specifies the study area and data. Section 3 describes the details of the proposed framework and each component. The experimental results and a discussion are provided in Section 4. Section 5 concludes the paper.

2. Study area and data

The experiments were conducted on a global-scale study site, as shown in Fig. 2(a) and two local-scale study sites, i.e., Beirut in Lebanon and Bata in Equatorial Guinea, as shown in Fig. 2(b) for a wide-scale statistical evaluation and actual application evaluation, respectively. In the global-scale study site, the xView2 Building Damage Assessment (xBD) dataset (Gupta et al., 2019a) was used for the model training and evaluation, and also to conduct a comprehensive ablation study for each

Table 1
The 19 disaster events selected in the xBD dataset.

Disaster type	Disaster event	Continent	Event date
Earthquake	Mexico City earthquake	North America	Sep 19, 2017
Wildfire	Portugal wildfires	Europe	Jun 17-24, 2017
Wildfire	Santa Rosa wildfires	North America	Oct 8-31, 2017
Wildfire	Carr wildfire	North America	Jul 23–Aug 30, 2018
Wildfire	Woolsey fire	North America	Nov 9–28, 2018
Wildfire	Pinery fire	Oceania	Nov 25–Dec 2, 2018
Volcano	Lower Puna volcanic eruption	Oceania	May 23–Aug 14, 2018
Volcano	Guatemala Fuego volcanic eruption	North America	Jun 3, 2018
Storm	Tuscaloosa, AL tornado	North America	Apr 27, 2011
Storm	Joplin, MO tornado	North America	May 22, 2011
Storm	Moore, OK tornado	North America	May 20, 2013
Storm	Hurricane Matthew	North America	Sep 28–Oct 10, 2016
Storm	Hurricane Florence	North America	Sep 10-19, 2018
Flooding	Monsoon in Nepal, India, Bangladesh	Asia	Jul–Sep, 2017
Flooding	Hurricane Harvey	North America	Aug 17–Sep 2, 2017
Flooding, Storm	Hurricane Michael	North America	Oct 7-16, 2018
Flooding	Midwest US floods	North America	Jan 3–May 31, 2019
Tsunami	Indonesia tsunami	Asia	Sep 18, 2018
Tsunami	Sunda Strait tsunami	Asia	Dec 22, 2018

component of the proposed model. In the two local-scale study sites, datasets of the Beirut port explosion and the Bata military barracks explosion were used to further verify the effectiveness of the proposed framework in real-world disaster response scenarios.

2.1. The global-scale study-site

A global scale study site was selected for the general building damage assessment. The locations of the sudden-onset disasters are shown in Fig. 2(a). These disasters mainly occurred in North America, Asia, and Australia, between 2011 and 2019. The details of these disaster events are presented in listed in Table 1. The disaster types include earthquakes, fires, volcanoes, storms, flooding, and tsunamis. Beyond the previous application scenario of a single disaster and a local area, these study areas represent a real-world scenario with abundant disaster types and a large spatio-temporal span, which can be used to evaluate the generalization ability of the proposed framework.

2.1.1. The xView2 building damage assessment (xBD) dataset

The xBD dataset contains 22,068 HSR bitemporal optical satellite images collected as part of the Maxar/DigitalGlobe Open Data Program (https://www.digitalglobe.com/ecosystem/open-data), covering a total of 45,361.79 km^2 and with 850,736 building instances. These optical satellite images have various spatial resolutions because they were collocated from different satellite platforms, e.g., WorldView-2 and WorldView-3. To assess the building damage in the multiple disaster types, we use the Joint Damage Scale, which was created with the help of the National Aeronautics and Space Administration (NASA), the California Department of Forestry and Fire Protection (CAL FIRE), the Federal Emergency Management Agency (FEMA), and the California Air National Guard (Gupta et al., 2019b), and is based on the HAZUS natural hazard analysis tool (Vickery et al., 2006), the Kelman scale (Kelman, 2003), and the EMS-98 scale (Grünthal, 1998). The Joint Damage Scale includes four discrete damage levels: *Non-Damage*, *Minor Damage*, *Major Damage*, *Destroyed*, as the damage classification criteria. A detailed description of each damage level is provided in Table 2. There are 313, 033, 36,860, 29,904, and 31,560 building instances, respectively, for the *Non-Damage*, *Minor Damage*, *Major Damage*, and *Destroyed*. The rest of the building instances are unclassified because the annotators were unable to identify them (Fig. 3).

Table 2
Joint Damage Scale description.

Damage level	Description
Non-damage	Undisturbed. No sign of water, structural or shingle damage or burn marks.
Minor damage	Building partially burnt, water surrounding structure, volcanic flow nearby, roof elements missing, or visible cracks.
Major damage	Partial wall or roof collapse, encroaching volcanic flow or surrounded by water/mud.
Destroyed	Scorched, completely collapsed, partially/completely covered with water/mud, or otherwise no longer present.

2.1.2. xBD dataset split

The xBD dataset is officially split into three parts: train, test, and holdout with a split ratio of 80%/10%/10%. The training set contains 18,336 images with 632,228 building polygons, while test set contains 1866 images with 109,724 building polygons and hold-out set contains 1866 images with 108,784 building polygons. Details of the xBD dataset are listed in Table 3. We followed the common practice and used the train/test/holdout sets for the training, ablation study, and benchmarking respectively.

2.2. Study sites for the local-scale man-made disasters

In the real world, sudden onset major disasters have diverse categories and are unpredictable, which means that a robust building damage assessment model should have a generalization ability for the out-of-distribution disasters that can happen anywhere. Local-scale man-made disasters meet this requirement. Thus, we chose two recent explosion events, Beirut port explosion and the Bata military barracks explosion, as examples to test the effectiveness of the proposed framework. These two study sites are shown in Fig. 2(b). The images of these two datasets are shared by Maxar/DigitalGlobe Open Data Program. Note that these two datasets were only used as test sets to evaluate the generalization of the trained model.

2.2.1. The Beirut port explosion

Beirut is the largest city and the capital of Lebanon, and is situated on a peninsula at the midpoint of Lebanon's Mediterranean coast. The Beirut port explosion event happened on August 4, 2020. It was caused by the accidental detonation of 2750 metric tons of ammonium nitrate, which is a common industrial chemical used in fertilizer, and is also a component of mining explosives. As of August 5, 2020, there were at least 135 people killed, around 5000 people injured and 300,000+ people displaced due to the extent of the damage (https://www.maxar.com/open-data/beirut-explosion).

The pre- and post-disaster images were collected by the WorldView-2 satellite, and were obtained on July 31, 2020, and August 5, 2020, respectively. These images have a spatial resolution of nearly 0.5 m and have off-nadir angles of 17.2° and 32.7°, respectively. These images have been preprocessed, including orthorectification, atmospheric compensation, dynamic range adjustment, and pan-sharpening. We cropped the images using the administrative boundary vector data of the

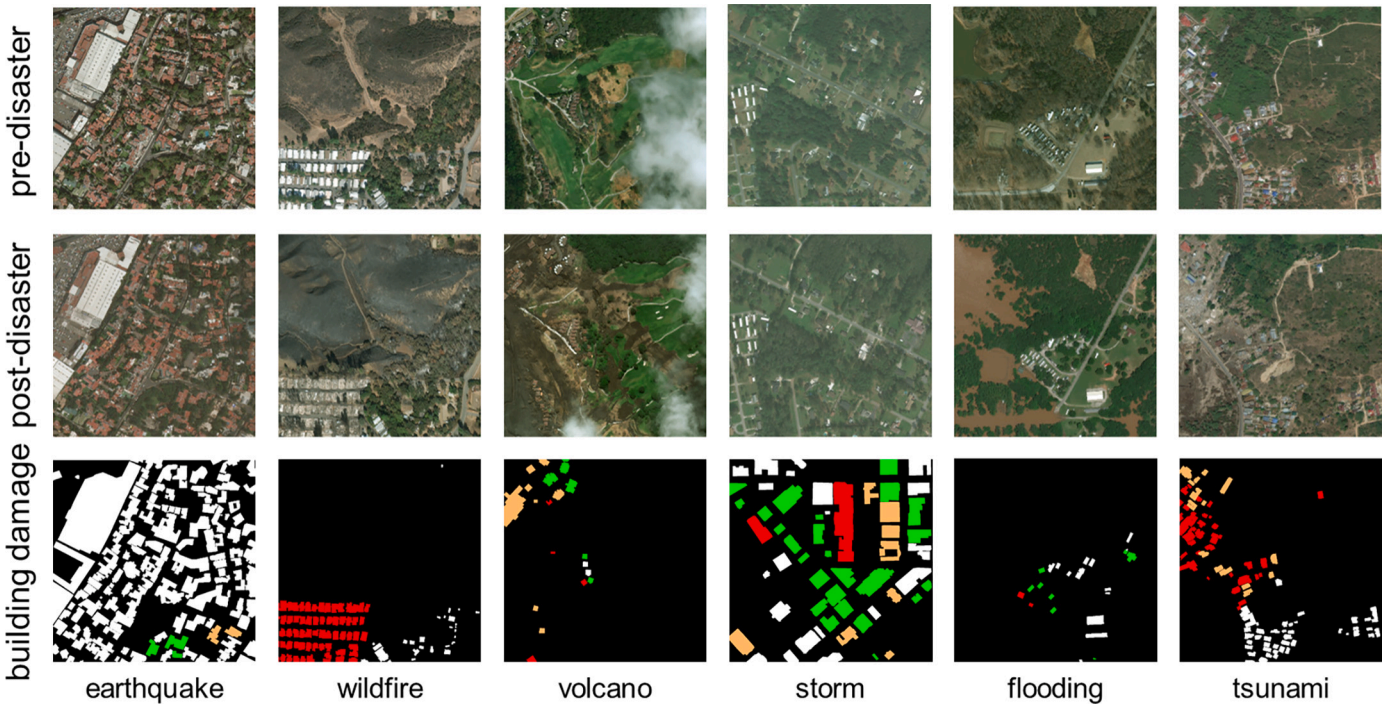

Fig. 3. Examples from the xBD Dataset.

Table 3
Statistics of the xBD dataset and the dataset split.

Split	#Images	#Building polygons	Usage
Train	18,336	632,228	Training
Test	1866	109,724	Ablation study
Holdout	1866	108,784	Benchmarking

whole of the city of Beirut. Each image covers an effective area of 19.8 km^2 with 16,744 × 11,880 pixels. To quantitatively evaluate the proposed framework, we recorded building polygons annotations and the damage state, which totaled 16,063 instances.

2.2.2. The Bata military barracks explosion

Bata is the largest city and the commercial capital of the Central African country of Equatorial Guinea. The Bata explosion event happened on March 7, 2021, killing at least 105 people and injuring more than 600 (https://www.maxar.com/open-data/bata-explosions). Almost all the buildings and homes in the city suffered huge damage. Officials claim that this event was caused by poorly stored explosives that detonated after nearby farmers conducted stubble burning.

The pre- and post- disaster images were collected by the GeoEye-1 satellite, on August 07, 2020, and March 09, 2021, respectively. These images have a spatial resolution of nearly 0.5 m and the off-nadir angles are 27.9° and 33.8°, respectively. We chose a sub-region of the city of Bata to demonstrate the effectiveness of the proposed framework. Each image covers nearly 16 km^2 with 10,033 × 8085 pixels. A total of 243 structures appear to have either been "heavily damaged or completely destroyed", according to a preliminary analysis by the United Nations Institute for Training and Research (https://unitar.org/maps/map/3258). We chose these data and some expert annotations to comprehensively evaluate the performance of the proposed framework in the annotation of large-scale man-made disaster events.

3. Methodology

3.1. Problem statement: building damage assessment

Building instance damage assessment is a hybrid task, made up of two subproblems: (1) building localization and (2) damage classification. The inputs and outputs of this task are shown in Fig. 4. The inputs are coregistered bitemporal pre- and post-disaster images. The outputs are a binary mask indicating the building position, and a multi-class mask indicating the degree of building damage.

3.1.1. Subproblem 1: building localization

The first step of building damage assessment is the building localization, which involves locating the building positions on the pre-disaster image as a reference. This task takes the pre-disaster image as input and outputs a binary mask, where the numbers 1 and 0 respectively represent whether a building exists or not.

3.1.2. Subproblem 2: damage classification

After establishing the building positions in the pre-disaster image, the degree of damage needs to be estimated for each building. Note that each building only has one degree of damage, as shown in Fig. 4(d). This means that this task outputs a multi-class mask based on the binary mask from **Subproblem 1**, where each pixel value indicates the degree of damage, and all the pixels of each building are semantically consistent.

3.1.3. The relationship with semantic change detection

Building damage assessment is highly relevant to semantic change detection, and can be seen as a special case of general semantic change detection. If we assume that there is an n-dimensional semantic category space for pre-change state and post-change state, then the general semantic change detection method needs to detect n^2 categories of change. For building damage assessment, there is a one-dimensional semantic category space for the pre-change state and an n-dimensional semantic category space for the post-change state. Only n categories of change therefore need to be detected, which significantly simplifies the semantic change detection problem. Therefore, we refer to this special

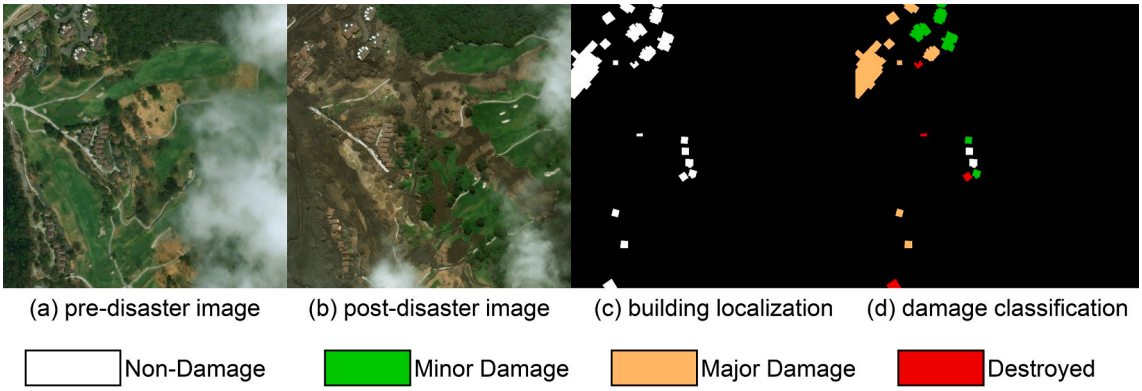

Fig. 4. The building damage assessment requires the inputs (a) pre-disaster image and (b) the post-disaster image, and then outputs the results of (c) the building localization and (d) the damage classification.

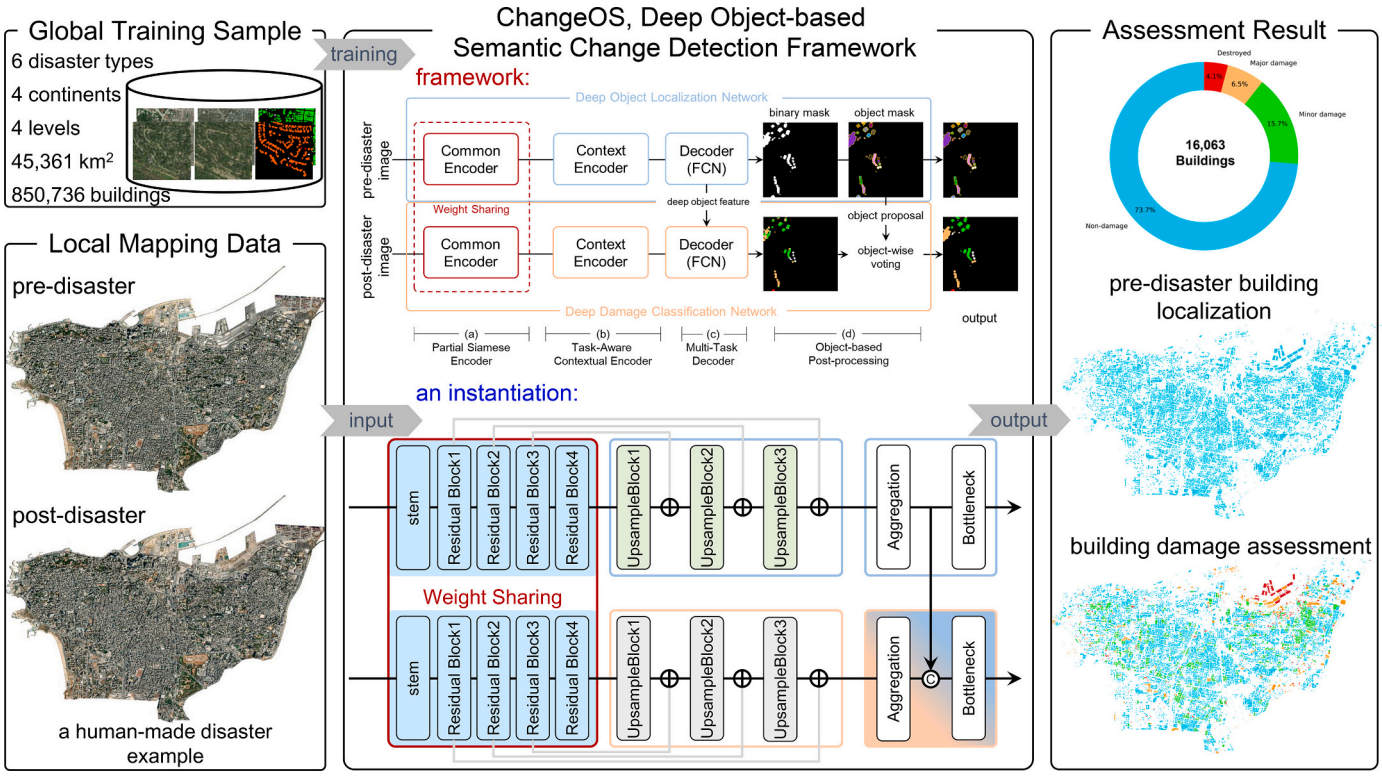

Fig. 5. Overview of ChangeOS framework. The key components are (a) partial siamese encoder, (b) task-aware contextual encoder, (c) multi-task decoder, and (d) object-based post-processing. ChangeOS directly takes bitemporal images as input, and outputs instance-level building damage assessment result, including the position and damage state of each building.

case as the *one-to-many semantic change detection problem*.

3.2. ChangeOS: the deep object-based semantic change detection framework

The building damage assessment can be seen as a one-to-many semantic change detection problem. Based on this insight, we propose the deep object-based semantic change detection framework (ChangeOS) for building damage assessment in a rapid global disaster response scenario using HSR bitemporal pre- and post-disaster images. The general design of ChangeOS follows the idea of OBIA, which features an object generation module and an object classification module. The whole framework consists of a deep object localization network as the object generation module and a deep damage classification network as the object classification module at the macro level, as shown in Fig. 5, which is constructed by four key components: (1) partial Siamese encoder, (2) task-aware contextual encoder, (3) multi-task decoder, and (4) object-based post-processing at the micro level. Compared with traditional object-based approach, ChangeOS is a deep object-based approach, which features (1) **deep object generation**. Deep object localization network which is a bottom-up instance segmentation method, is proposed to generate objects in place of super-pixel segmentation; (2) **end-to-end training and inference**. The object generation module is integrated with the object classification module in a differentiable manner via deep object features; (3) **consistent semantic within an object**. The object classification is semantically consistent within an object.

Differing from the conventional Siamese encoder design, we adopt a partial Siamese encoder design. This is because we believe that different tasks need different spatial context modeling. We use the partial Siamese

encoder to extract task-independent deep features and then use the task-aware contextual encoder to further extract the task-aware context to enhance the deep features. For multi-task prediction, we adopt a multi-task decoder and achieve multi-task feature interaction. Thus, the deep object features can guide the damage classification to achieve feature-level interaction. To further guarantee the semantic consistency of the object, we adopt object-based post-processing, which takes each object as the basic classification unit to adjust the predicted category of pixels inside an object via simple voting. These components are described in detail in the next sub-sections.

3.2.1. The fully convolutional neural network (FCN)

The fully convolutional neural network (FCN) (Long et al., 2015) is a variant of the convolutional neural network (CNN), and is widely used in semantic image segmentation. The FCN is the fundamental building component of the proposed ChangeOS framework. The CNN architecture is always stacked by multiple convolutional layers followed by non-linear activation and normalization, and follows a multi-stage design where each stage outputs a different resolution feature map. As the CNN goes deeper, the resolution of the feature maps gets coarser but with stronger semantic information (Ronneberger et al., 2015; Lin et al., 2017). CNNs, such as AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), and ResNet (He et al., 2016) have been specifically designed for image classification.

To apply the CNN to a dense prediction task (e.g., semantic segmentation), the FCN removes the last fully connected layer in the CNN. This makes the FCN output feature maps with spatial layout, which is essential to a dense prediction task. In addition, the FCN architecture usually introduces the upsampling module to recover the resolution of the feature maps for pixel-wise prediction. The upsampling module has many candidates, such as a deconvolutional layer (Noh et al., 2015), bilinear upsampling followed by a convolutional layer (Chen et al., 2018), nearest-neighbor upsampling followed by a convolutional layer (Lin et al., 2017), etc. To optimize the FCN for semantic image segmentation, pixel-wise cross-entropy loss is commonly used. More details can be found in Long et al. (2015). The proposed method belongs to the FCN-family.

3.2.2. The partial Siamese encoder

The partial Siamese encoder is an FCN-family model used to extract task-independent deep features for downstream tasks. For bitemporal optical images, we introduce a weight sharing mechanism to reuse the network architecture and its weight to alleviate overfitting problem. This is motivated by the fact that the bitemporal images belong to the same modality and have similar visual patterns since the pre- and post-disaster images were collected by the same optical sensor, but in different temporal phases. In this case, two independent network weights would produce huge and redundant parameter space, impeding the network training. Therefore, sharing the weights can significantly reduce the complexity of the parameter space to alleviate the overfitting problem.

We implement the partial Siamese encoder based on ResNet (He et al., 2016). The vanilla ResNet is used for the image classification. By only keeping a stem block and four residual blocks, ResNet can be used as a hierarchical feature extractor. The stem block is made up of a 7×7 convolutional layer with stride 2 followed by batch normalization (BN), rectified linear unit (ReLU) activation, and a max-pooling layer. The residual block is made up of many residual units. For example, the 50-layer ResNet has 3, 4, 6, and 3 residual units for the four residual blocks, respectively. Each residual unit is made up of stacked 1×1, 3×3 and 1×1 convolutional layers, where each convolutional layer is followed by BN and ReLU. A shortcut connection is also applied to the input and the output to solve the vanishing gradient problem. Taking a single-temporal image as input, the feature maps from the four residual blocks of ResNet are extracted for the downstream tasks. For bitemporal images, this forward computation is repeated twice to obtain bitemporal feature maps.

3.2.3. The task-aware contextual encoder

For different tasks, it is necessary to capture contextual information from different ranges. To this end, we propose a task-aware contextual encoder to further extract task-aware contextual information to obtain task-aware and context-enhanced deep features. The task-aware contextual encoder is made up of three upsample blocks, and for building localization and damage classification, the network architectures are identical but with different weights, as shown in Fig. 5(b). Similar to the feature pyramid network (FPN) (Lin et al., 2017), we also introduce a lateral connection to combine the high-level feature maps with longer range context and the low-level feature maps with finer spatial details. To align the spatial resolution of the low-level and high-level feature maps, the upsample block is designed, which is a nearest-neighbor upsampling layer with a scale factor of 2, followed by a 1×1 convolutional layer. The pointwise addition is simply used to fuse the two feature maps. In this way, four task-aware and context-enhanced feature maps can be obtained for the subsequent multi-task decoding.

3.2.4. The multi-task decoder

The building localization and damage classification tasks can be both seen as semantic segmentation tasks in ChangeOS. For further object-based classification, we improve the building localization task from semantic segmentation to instance segmentation by a connected component labeling algorithm (Wu et al., 2005), which outputs a set of object polygons but can be supervised by a pixel-wise loss function. The multi-task decoder takes feature maps of four scales as the input and outputs a binary probability map to indicate the position of the buildings and a multi-class probability map to indicate the damage classification probability for each pixel. This multi-task decoder is made up of two identical sub-networks.

3.2.4.1. The object localization sub-network. A simple FCN is adopted to perform binary segmentation, which is made up of an aggregation module and a bottleneck block (He et al., 2016). The aggregation module takes feature maps of four scales as input, which have output strides[1] of 4, 8, 16, and 32, respectively. This module progressively applies $2\times$ upsampling to each feature map until the feature maps all have an output stride of 4, where the upsampling operation is implemented by a bilinear upsampling layer and a 3×3 convolutional layer. We then sum over all the feature maps to produce a multi-scale fused feature map with the output stride of 4 and feature channels of 256. This multi-scale fused feature map is also forwarded into the damage classification sub-network as deep object feature, providing the object prior. The bottleneck block is made up of three convolutional layers, each followed by batch normalization and ReLU. The kernel size of the middle convolutional layer is 3×3, while the other two are 1×1. Meanwhile, identity mapping is applied to construct a residual connection. A final 1×1 convolutional layer with one output channel is attached to produce the binary probability map, which is directly supervised by the localization loss L_{loc}. The final building localization map is obtained by $4\times$ bilinear upsampling.

3.2.4.2. The damage classification sub-network. An identity FCN is used for the damage classification. In addition, the deep object feature from the object localization sub-network is concatenated with the classification feature to provide the object prior for modeling the temporal difference. In this way, at the macro level, the deep object localization network is integrated with the deep damage classification network in a differentiable manner, thus achieving end-to-end training and inference.

[1] The output stride is defined as the ratio of the input image size to the output feature map size.

As the classification layer, a 1 × 1 convolutional layer with five output channels is attached, which is supervised by the classification loss L_{cls}. The final damage classification result is obtained by 4× bilinear upsampling and the argmax operation.

3.2.4.3. The multi-task loss function for joint optimization. To jointly optimize the building localization and damage segmentation, we designed a multi-task loss function. The overall multi-task loss function is formulated as follows:

$$L = L_{\text{loc}} + L_{\text{cls}} \tag{1}$$

The commonly used binary segmentation loss is a combination of binary cross-entropy and dice loss (Milletari et al., 2016). However, in this task, guaranteeing the recall of building localization is very important than precision, because it is a prefix task of damage classification. Therefore, dice loss is not optimal choice in this scenario. To this end, we replace the dice loss with Tversky loss based on the Tversky index (Tversky, 1977), which can control the trade-off between recall and precision by two hyperparameters. Thus, our building localization loss L_{loc} is the sum of the binary cross-entropy loss and Tversky loss:

$$\begin{aligned} L_{\text{loc}}(p, y, \alpha, \beta) = & \frac{1}{N} L_{\text{bce}}(p, y) + L_{\text{tver}}(p, y, \alpha, \beta) \\ = & -\frac{1}{N}(y\log(p) + (1-y)\log(1-p)) \\ & +1 - \frac{\sum_{i=1}^{N} p_i y_i}{\sum_{i=1}^{N} p_i y_i + \alpha \sum_{i=1}^{N} (1-p_i) y_i + \beta \sum_{i=1}^{N} p_i (1-y_i)} \\ & 0 \le \alpha, \beta \le 1 \end{aligned} \tag{2}$$

where p and y denote the predicted class probability and ground truth. α and β are hyperparameters introduced by the Tversky loss function to control the penalizing ratio for the false negatives and false positives, respectively. We set α to 0.9 and β to 0.1, because the recall of the building localization is more significant than precision in building damage assessment.

For the damage classification loss L_{cls}, we use multi-class cross-entropy loss:

$$L_{\text{cls}}(p, y) = -\frac{1}{N} y\log(p) \tag{3}$$

3.2.5. Object-based post-processing

The building localization and damage classification obtained by the multi-task decoder are both the pixel-level classification results. However, this pixel-level representation always causes partial damage recognition. For example, fire can produce burn marks on a building, and the burn marks may partially appear on the building roof, which results in that only the burn marks being recognized as minor damage, while the rest of the building is identified as no damage. The intrinsic reason for this lies in the fact that the damage degree refers to the property, and thus belongs to a building object rather than a pixel. This means that all the pixels of a building object should be semantically consistent.

To adjust these semantically inconsistent building objects, we further propose an object-based post-processing method in the ChangeOS framework to make the building instance semantically consistent, following OBIA (Blaschke, 2010). The object-based postprocessing includes two steps: (1) object proposal, and (2) object-wise weighted voting, as shown in Fig. 5(d). The object proposal involves extracting each building object in the building localization result, using a connected component labeling algorithm (Wu et al., 2005). The object-wise weighted voting algorithm then collects all the pixels of each building object and weights these pixels by each class to obtain a unique damage degree for each building object.

4. Experiments

4.1. Experimental settings

4.1.1. Implementation details

All the models were trained for 60k iterations using stochastic gradient descent (SGD) with a "poly" learning rate policy, where the initial learning rate was set to 0.03 and was then multiplied by $(1 - \frac{1}{1-\text{max\_i ter}})^{\gamma}$ with γ of 0.9. The weight decay was set to 0.0001 and the momentum was set to 0.9 for the SGD. The batch size totaled 16 over two Titan RTX GPUs and synchronized batch normalization was adopted to obtain more stable statistics. Horizontal and vertical flip and random rotation of $90^{\circ} \cdot k$ were used for the training data augmentation.

4.1.2. Metrics

We use the standard xView2 metric to evaluate the results of the building damage assessment methods. The xView2 metric is a variant of the F_1 score, jointly considering the localization and damage classification performance. The standard F_1 score can be computed as follows:

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{4}$$

where TP, FP and FN are the numbers of true positive, false positive and false negative pixels, respectively.

To evaluate the localization quality, the localization score F_1^{loc} is defined as a standard pixel-based F_1 score for binary classification. The damage classification needs to classify the pixels into one of four classes: non-damage, minor damage, major damage, and destroyed. To evaluate the damage classification, for each class, a standard pixel-based F_1 score is used. The damage classification score F_1^{dam} is then computed by their harmonic mean, as follows:

$$F_1^{\text{dam}} = \frac{4}{\frac{1}{F_1^{\text{no dmg.}}} + \frac{1}{F_1^{\text{minor dmg.}}} + \frac{1}{F_1^{\text{major dmg.}}} + \frac{1}{F_1^{\text{destroyed}}}} \tag{5}$$

The overall xView2 metric can be computed by the weighted sum of the localization score and damage classification score, as follows:

$$F_1^{\text{overall}} = 0.3F_1^{\text{loc}} + 0.7F_1^{\text{dam}} \tag{6}$$

4.2. Benchmark methods

To evaluate the proposed framework, we adopted the xView2 baseline method (https://github.com/DIUx-xView/xView2_baseline) and the xView2 1st place solution method (https://github.com/DIUx-xView/xView2_first_place) as a comparison. Because the 1st place solution is based on a multi-model ensemble, for a fair comparison, we used the ResNet-34 based models as a comparison.

1. UNet + ResNet, xView2 baseline. This approach follows a cascade pipeline of first pixel-wise locating the buildings and then patch-wise classifying the damage-level of the buildings. It also belongs to an object-based method because its basic classification unit is an object. UNet was used for the building localization and ResNet was used for the damage classification. These two models were respectively trained by pre-disaster images and post-disaster images.
2. Siamese-UNet, xView2 1st place solution (four models). This approach adopts the UNet architecture for both the pixel-wise building localization and damage classification. The building localization UNet was first trained on pre-disaster images. The damage classification UNet was then trained on pre- and post-disaster image pairs. To overcome the knowledge gap, the damage classification UNet was initialized by the parameters of the building localization UNet before training, namely Siamese-UNet. This approach adopts a

multi-model ensemble strategy to improve the final performance. One building localization UNet and three damage classification UNets were used in the ensemble.

4.2.1. Relation to xView2 1st place solution

Here, we discuss the relations between the proposed method and Siamese-UNet, xView2 1st place solution from three perspectives: (1) Overall framework. xView2 1st place solution follows pixel-based framework, while the proposed method follows object-based framework. (2) Network architecture. xView2 1st place solution uses a UNet and a Siamese-UNet architecture for building localization and damage classification. We designed a unified partial Siamese and task-aware FPN-like model for joint building localization and damage classification. (3) Training strategy. xView2 1st place solution adopts a two-stage training pipeline, which first trains a UNet for building localization and then trains a Siamese-UNet for damage classification, while we adopt end-to-end training, which is more friendly to many real-world applications.

4.3. Benchmark comparison and analysis

The benchmark results were obtained by evaluation on the xView2 `holdout` split. The results are listed in Table 4. ChangeOS with ResNet-50 achieves the best performance and significantly outperforms the other two approaches by means of the systematic technical improvement. Even when using a shallower feature extractor, such as ResNet-18 or 34, ChangeOS still achieves extraordinary performance. For a fair comparison, we also evaluated the pixel-based ChangeOS, which is the standard ChangeOS without object-based post-processing. The results suggest that although without object-based post-processing, ChangeOS still can achieve competitive performances when using different backbones, e.g., with the same backbone of ResNet-34, pixel-based ChangeOS achieves 74.743% $F_1^{overall}$ score, which is 3.05% higher than the Siamese-UNet. This confirms the superior of the proposed network design to other compared methods. We also benchmarked the inference speed. UNet+ResNet needs 14.26 s on the CPU, while the ChangeOS only needs 3.97–5.65 s on the CPU for a pair of pre- and post-disaster images. Moreover, ChangeOS features end-to-end inference, which makes it more suitable for GPU acceleration. With further GPU acceleration, ChangeOS achieves a sub-second inference speed and only needs nearly 0.5 s on the GPU. Although the Siamese-UNet approach can also utilize GPU acceleration, it still costs 4.43 s on the GPU because of the multi-model ensemble and two-stage pipeline. This suggests that ChangeOS fundamentally improves the pipeline of building damage assessment, making the process faster and more accurate. These characteristics make it possible to achieve automatic and rapid building damage assessment for rapid disaster response.

To explain the error source of the benchmarked methods, we present their confusion matrices in Fig. 6. It can be seen that UNet + ResNet fails to recognize the minor damage and major damage, which is the most common error source. Because UNet + ResNet models the damage classification as an image classification task, this local vector representation has difficulty in characterizing the details of the building damage and long-range disaster context. To break this limitation, Siamese-UNet and ChangeOS adopt a fundamentally different feature representation, i.e., pixel-wise representation, which models the damage classification as a semantic segmentation task. By the pixel-wise representation, the building damage details can be more easily obtained for further analysis. From Fig. 6(b), it can be seen that Siamese-UNet achieves a reasonable classification performance over the four damage levels, but it achieves a poor F_1 scores when compared with ChangeOS. This suggests that Siamese-UNet tends to recognize a large number of pixels as background, which causes the relatively low recalls of the four damage classes. This is because Siamese-UNet uses a full weight-sharing encoder, which is unable to model specific context information for different tasks. However, minor damage recognition is still challenging, in that ChangeOS tends to recognize minor damage as no damage or major damage. This is because most minor damage belongs to partial building damage, such as a partially burnt building. The partially damaged area becomes the key factor in the damage classification, which is a potential roadmap for further improvement.

The visual performance of building damage assessment is equally important because the assessment map product can reflect the distribution of the different damage-level buildings. This is important guidance for humanitarian assistance and disaster recovery. The visual results are shown in Fig. 7, which shows that ChangeOS can produce a more accurate and semantically consistent assessment map to reflect the damage level distribution due to the introduction of deep OBIA. For example, the tsunami resulted in many destroyed buildings and many buildings with major damage, but UNet-ResNet can guarantee semantic consistency for each object because it follows the OBIA framework. However UNet-ResNet is unable to accurately recognize buildings with major damage. Siamese-UNet can recognize some of the buildings with major damage, but without semantic consistency for each building. Benefiting from the deep OBIA framework, ChangeOS can accurately recognize these buildings and guarantee semantic consistency.

4.4. Ablation study

ChangeOS has multiple technical improvements, including: (1) end-to-end multi-task training and inference; and (2) the deep OBIA (deep object feature, object-based post-processing). To delve into ChangeOS, an ablation study was conducted to evaluate the contribution of the key

Table 4
Benchmark comparison on the xView2 `holdout` split. ChangeOS was evaluated with four different backbone networks. The inference time was recorded using an Intel (R) Xeon(R) CPU E5-2690 v4 @ 2.60 GHz for the CPU time, and a Titan RTX GPU for the GPU time. The *object-based* ChangeOS is the standard ChangeOS. The *pixel-based* ChangeOS is the standard ChangeOS without object-based post-processing.

Method	Backbone	$F_1^{overall}$(%)	F_1^{loc}(%)	F_1^{dam}(%)	Damage F_1 per class (%)				Inference time (s)	
					No Dmg.	Minor Dmg.	Major Dmg.	Destroyed	CPU	GPU
pixel-based										
Siamese-UNet	ResNet-34	71.683	85.917	65.583	86.736	50.025	64.432	71.679	–	4.43
ChangeOS	ResNet-18	74.296	84.618	69.872	88.612	52.097	70.364	79.650	2.13	0.36
ChangeOS	ResNet-34	74.743	85.157	70.281	88.631	52.379	71.163	80.079	2.48	0.36
ChangeOS	ResNet-50	75.238	85.413	70.877	88.979	53.326	71.239	80.597	3.13	0.37
ChangeOS	ResNet-101	75.502	85.693	71.135	89.113	53.113	72.445	80.788	3.58	0.37
object-based										
UNet + ResNet	ResNet-50	27.681	80.932	4.860	65.281	6.593	1.606	29.922	14.26	–
ChangeOS	ResNet-18	77.110	84.618	73.892	92.378	57.407	72.541	82.624	3.97	0.49
ChangeOS	ResNet-34	77.519	85.157	74.246	92.189	58.069	72.843	82.795	4.14	0.50
ChangeOS	ResNet-50	78.569	85.413	75.635	92.656	60.138	74.180	83.449	5.01	0.51
ChangeOS	ResNet-101	78.519	85.693	75.444	92.809	59.377	74.649	83.288	5.65	0.52

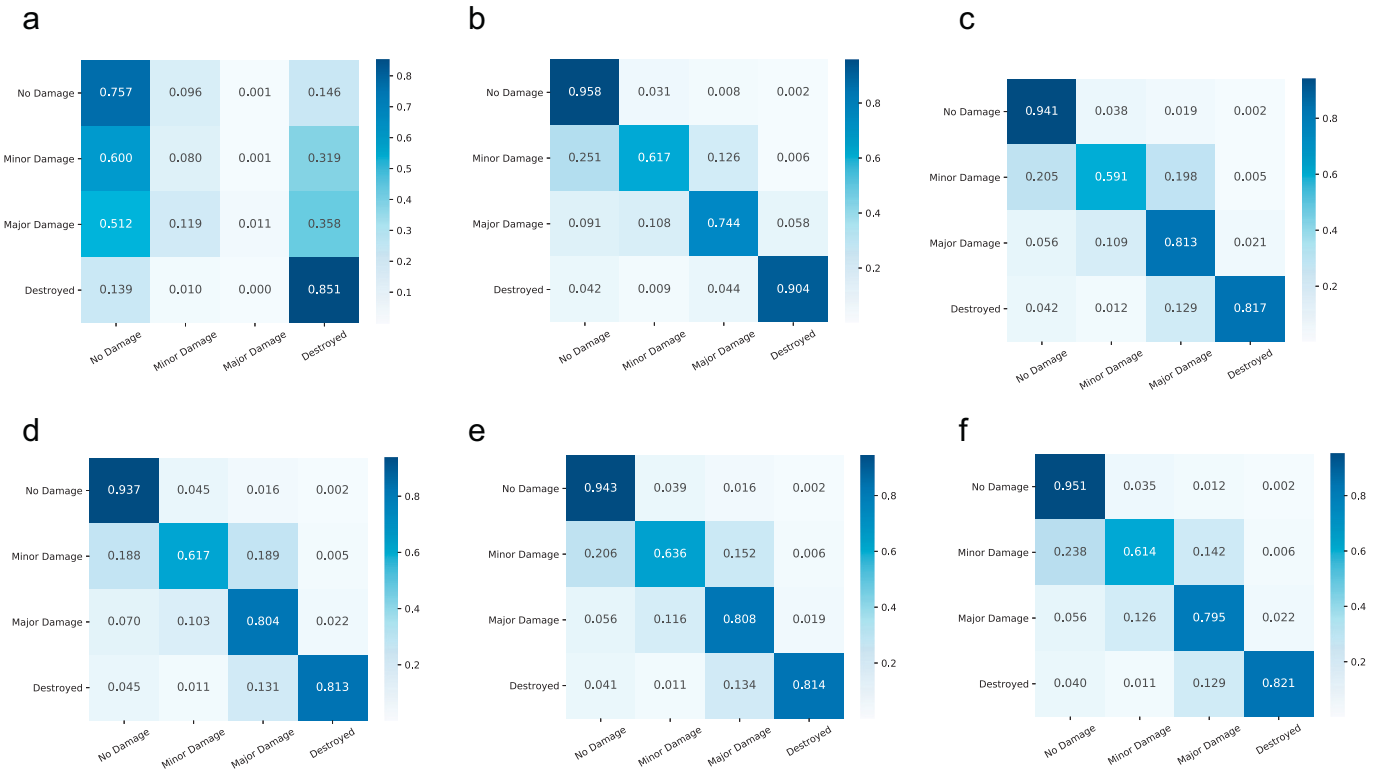

Fig. 6. Confusion matrices for the benchmarked methods. (a) UNet+ResNet. (b) Siamese-UNet. (c) ChangeOS with ResNet-18. (d) ChangeOS with ResNet-34. (e) ChangeOS with ResNet-50. (f) ChangeOS with ResNet-101.

components and the impact of the important hyperparameter of the loss function. We report these ablation studies by evaluation on the `test` split.

4.4.1. *End-to-end multi-task training and inference*

ChangeOS features end-to-end optimization and inference for building damage assessment, which is the most significant improvement at the system level. ChangeOS without deep object features and object-based post-processing can still achieve F_1^{overall} scores of 66–67% with different backbone networks. This suggests that the network architecture of ChangeOS is compatible with end-to-end multi-task learning.

4.4.2. *Deep object feature for damage classification*

To ensure the contribution of the deep object feature, we compared ChangeOS with and without deep object feature by adding or removing the concatenation of this feature in the multi-task decoder. The results presented in Table 5 suggest that the deep object features are very important for the damage classification, and they significantly improve the damage classification F_1^{dam} score by 10–12% for the four different backbone networks. Furthermore, there is little impact on the building localization F_1^{loc} score when these feature are forwarded from the localization sub-network to the damage classification sub-network to provide pre-disaster object prior.

4.4.3. *Object-based post-processing*

To investigate the object-based post-processing, we first evaluate ChangeOS with object-based post-processing but without deep object features. To our surprise, the F_1^{overall} score shows a similar performance to ChangeOS with only deep object features. This means that bitemporal features can provide an object-prior for damage classification to make building instance semantics consistent. However, a natural question arises: *whether is there a redundant object-prior from both bitemporal features and object proposal*? To further validate the compatibility between the deep object features and object-based post-processing, ChangeOS with both deep object features and object-based post-processing was evaluated. The results listed in Table 5 suggest that these two technical improvements are compatible. For example, ChangeOS with a backbone network of ResNet-101 achieves 78.759% F_1^{overall} score, where the deep object features bring an extra 3.378% improvement. Meanwhile, there is an interesting phenomenon that ChangeOS with deeper backbone network obtains a greater improvement, which may be because the deeper backbone network can obtain a more accurate object-prior from the building localization sub-network. For the visual performance analysis, the intermediate results of ChangeOS are presented in Fig. 8, where it can be seen that the object-based post-processing can effectively alleviate the semantic inconsistency when the object proposal module accurately extracts the building instances.

4.4.4. *The hyperparameters in the Tversky loss*

The Tversky loss function introduces two hyperparameters: α and β, which control the trade-off between the precision and recall. In common practice, β is always equal to $1-\alpha$. Therefore, we only investigated the impact of hyperparameter α on the performance. We chose 10 different α values from 0.1 to 0.9 with an interval of 0.1 for analysis. The results are presented in Fig. 9. The larger α means that the trained model obtains a higher recall for the building localization sub-task. Intuitively, the F_1^{overall} score and damage classification F_1^{dam} score are highly linearly related to the value of α. As α increases, the building localization F_1^{loc} score first increases and then decreases, and it achieves an optimal trade-off between precision and recall when α is equal to 0.5. This means that higher recall for the building localization is beneficial to building damage assessment because building localization is an important pre-task for the damage classification. If a pixel is judged as true positive for the damage classification, it must be first predicted as a positive in the building localization. Therefore, the recall of the building localization sub-task is important for the building damage assessment. Meanwhile, the building damage classification sub-task has a larger weight of 0.7 in the overall score, which is biased to penalize the miss recognition, thereby

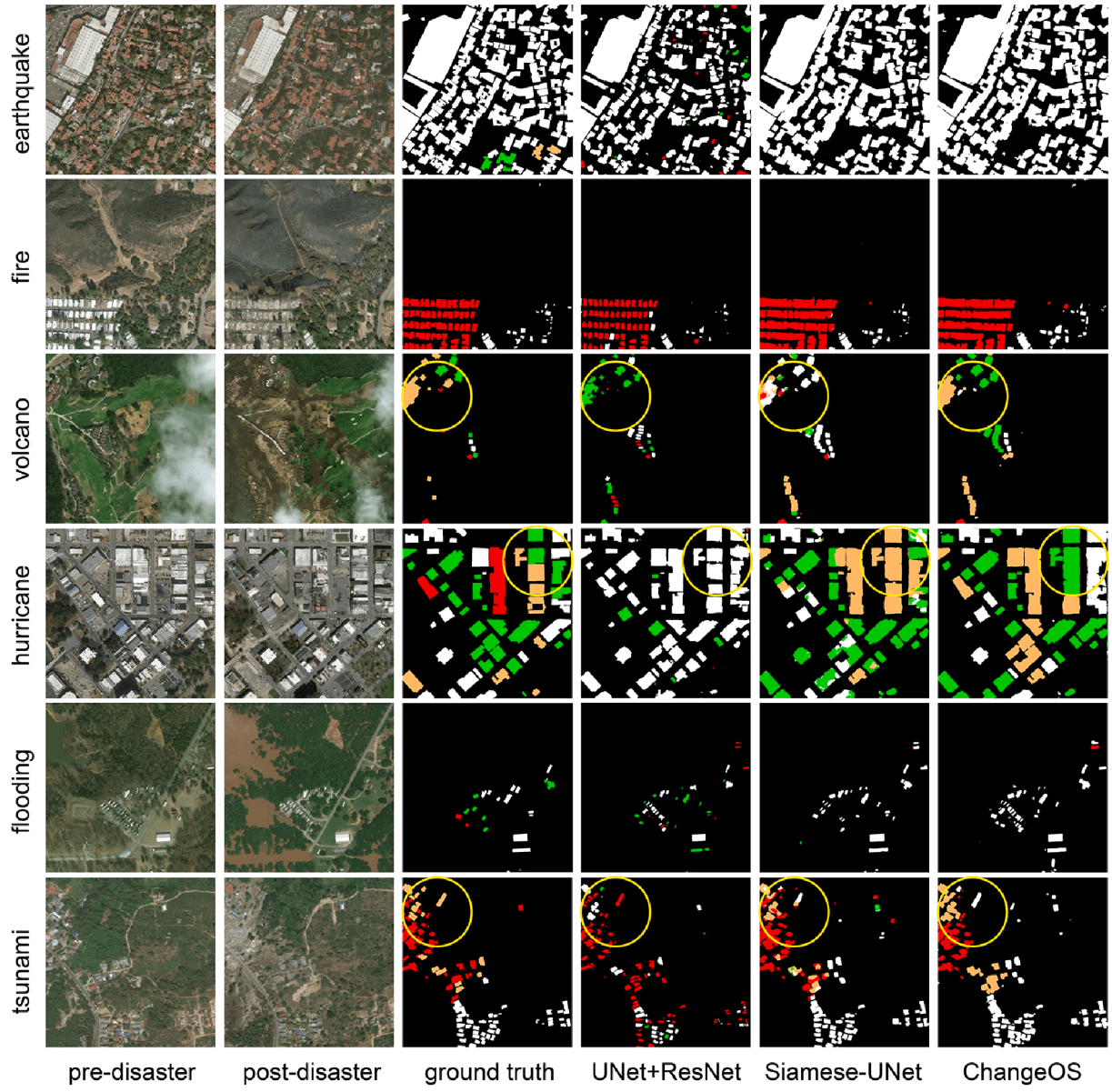

Fig. 7. Visual performance of the benchmark methods on a subset containing six types of disaster from the xBD dataset.

improving the importance of the recall of the building localization. Therefore, in ChangeOS, we use a α of 0.9 and a β of 0.1 as default hyperparameters in the Tversky loss.

4.5. Application for local-scale man-made disasters

In this section, we demonstrate the application of real-world building damage assessment for local-scale man-made disasters using the proposed method. We chose two recent explosion events, i.e., the Beirut port explosion and the Bata military barracks explosion, for the demonstration.

4.5.1. Experimental settings

4.5.1.1. Image preprocessing. These optical image products have been preprocessed, including orthorectification, atmospheric compensation, dynamic range adjustment, and pan-sharpening. For time-sensitive application scenarios such as rapid disaster response, there is no extra preprocessing applied. Therefore, there is always a misregistration problem for these relatively raw image pairs. This requires the model to handle the misregistration problem.

4.5.1.2. Model training and inference. We trained ChangeOS with ResNet-101 using the xBD training dataset. The implementation details were the same as the previous experimental settings. For the inference, we adopted a non-overlapping sliding window strategy to handle these images with very high resolutions and extensive coverage, due to the limited GPU memory. The window size was 512×512 and the stride was 512 pixels. To efficiently apply GPU acceleration, we adopted batch inference with a batch size of 4.

4.5.1.3. Accuracy assessment. To evaluate the accuracy of the whole pipeline, we computed the standard xView2 metric based on the manually annotated ground truth. The annotation of the ground truth was conducted by experts. In total, 16,063 and 5571 building instances were collected for the Beirut port explosion event and the Bata military barracks explosion event, respectively. Considering that, for many of the damaged buildings, it is difficult to recognize their damage level even for experts, we ignored these buildings during the evaluation. The trained model was directly evaluated on all the annotated samples

Table 5
Ablation study for the proposed modules. "Deep object features?" indicates whether deep object features are introduced from the localization sub-network into the damage classification sub-network. "Object-based?" indicates whether object-based post-processing is applied.

Backbone	Deep object feature?	Object-based?	$F_1^{overall}$(%)	F_1^{loc}(%)	F_1^{dam}(%)	Damage F_1(%) per class			
						No Dmg.	Minor Dmg.	Major Dmg.	Destroyed
ResNet-18			66.523	84.851	58.667	83.301	42.691	64.247	58.184
	✓		73.624	84.591	68.924	89.077	49.563	70.727	80.045
		✓	75.100	84.851	70.921	92.240	55.793	72.658	72.083
	✓	✓	76.953	84.591	73.680	92.855	55.665	73.701	83.431
ResNet-34			67.129	84.909	59.509	83.586	44.329	64.151	58.460
	✓		74.232	85.029	69.605	89.045	50.575	71.155	80.578
		✓	75.043	84.909	70.815	92.274	55.233	72.899	72.334
	✓	✓	77.072	85.029	73.663	92.679	55.276	73.682	84.399
ResNet-50			67.454	85.414	59.756	84.095	43.987	66.551	57.847
	✓		74.833	85.177	70.400	89.395	51.616	71.686	81.248
		✓	75.501	85.414	71.252	92.175	56.210	74.151	71.363
	✓	✓	77.558	85.177	74.293	92.994	56.208	74.420	84.324
ResNet-101			67.210	85.576	59.338	84.027	43.852	65.131	57.634
	✓		75.661	85.549	71.424	89.724	53.077	72.699	81.551
		✓	75.381	85.576	71.011	92.272	56.110	72.464	72.101
	✓	✓	78.759	85.549	75.849	93.493	58.862	75.493	84.711

without any fine-tuning, to consider the generalization ability for different spatial positions and disaster events.

4.5.2. Experimental results

4.5.2.1. Building damage assessment for the Beirut port explosion event. After the explosion, it was very dangerous to investigate the explosion and impacted area one the ground. We therefore used bitemporal pre- and post-disaster HSR remote sensing images and applied ChangeOS trained on the xBD dataset to safely and rapidly assess the building damage. The building damage assessment results for the Beirut port explosion event are shown in Fig. 10. Intuitively, the center of the explosion can be easily found in Fig. 10(d), because the cluster of destroyed buildings (rendered in red) is clearly the center of the explosion. This can help further the decision-making for the government in disaster response. For the explosion center, a detailed depiction damage state of the buildings was obtained as shown in Fig. 10(f).

Quantitatively, ChangeOS achieves a pre-disaster building localization F_1 score of 64.498% and a post-disaster building damage classification F_1 score of 46.860%, which is better than the other two compared approaches, as shown in Table 6. In particular, the destroyed and major damage building classification scores are superior to those of UNet+ResNet and Siamese-UNet. However, for the bitemporal images of the city of Beirut, the non-damaged building classification scores are much lower than the results obtained on the xView2 holdout set. A potential reason for this is that the large off-nadir angle makes the tall urban buildings appear more inclined in the image, thereby causing a more serious misregistration problem. For the misregistration problem, ChangeOS mainly benefits from the deep OBIA framework. Although the conventional OBIA approach has recently confirmed its effectiveness with regard to the misregistration problem (Liu et al., 2021), the weaker feature representation ability in conventional OBIA is shown in the building damage assessment accuracy of UNet + ResNet. In this scenario, Siamese-UNet with stronger feature representation achieves a better accuracy, although it is not an OBIA framework, which suggests that strong feature representation is important for damage classification. ChangeOS seamlessly combines OBIA and strong feature representation, i.e., deep OBIA, achieving the best accuracy among different methods. This helps to overcome the weak feature representation problem in the conventional OBIA framework.

In addition to a high accuracy, the inference speed is also important in disaster response. UNet + ResNet took 1978s on the CPU and Siamese-UNet needed 927s on a Titan RTX GPU. Under the same hardware environment, ChangeOS only required 52s on the GPU for building damage assessment of the whole of the city of Beirut, which covers 19.8 km^2, with the pre- and post-disaster images both having a size of 11,880 × 16,744 pixels. This suggests that ChangeOS can provide rapid and robust building damage assessment results for use in disaster response.

4.5.2.2. Building damage assessment for the Bata military barracks explosion event. The building damage assessment for the Bata military barracks explosion event is shown in Fig. 11. As with the Beirut port explosion event, the Bata military barracks explosion event also has an obvious disaster center, which can also be easily observed in the obtained our mapping product, as shown in Fig. 11(b). However, differing from the Beirut port explosion, the impacted area for the Bata military barracks explosion is smaller, but the buildings in the impacted area are almost destroyed, as shown in Fig. 11(b), (d) and (f). The quantitative performance is provided in Table 7. An important difference with the Beirut port explosion is that the building localization accuracy and non-damaged building classification accuracy are much higher in the Bata military barracks explosion event dataset. This is because the disaster center is not in a developed urban scenario. The buildings in this area mostly belong to low-rise buildings. Therefore, the bitemporal images with large off-nadir angles have little impact on the assessment results. These quantitative results further confirm the superiority of ChangeOS. This suggests that the seamless combination of OBIA and deep learning is a solid foundation for the change detection problem. Furthermore, the advantage of the inference speed of ChangeOS remains significant in the Bata military barracks explosion event dataset. ChangeOS only took 23 s on the GPU to complete the whole assessment procedure, while Siamese-UNet cost 342s on the same hardware environment.

4.5.2.3. Damaged statistic buildings for the two events. Based on the assessment products for the two man-made disaster events, we show the statistics of the damaged buildings in Fig. 12. We collected 16,063 building polygons and 5571 building polygons for the Beirut port explosion event and Bata military barracks explosion events. The common feature of these two explosion events was that the study areas are both locally impacted, which can be observed by their small ratios of the number of damaged buildings to the total. The difference lies in the fact that the cause of the Beirut Explosion was 2750 metric tons of ammonium nitrate, so the impacted area is obviously larger than Bata military barracks explosion. This also caused that there are larger other damage ratios in the Beirut port explosion event. Furthermore, these industrial chemicals will also have a potential impact on the surrounding

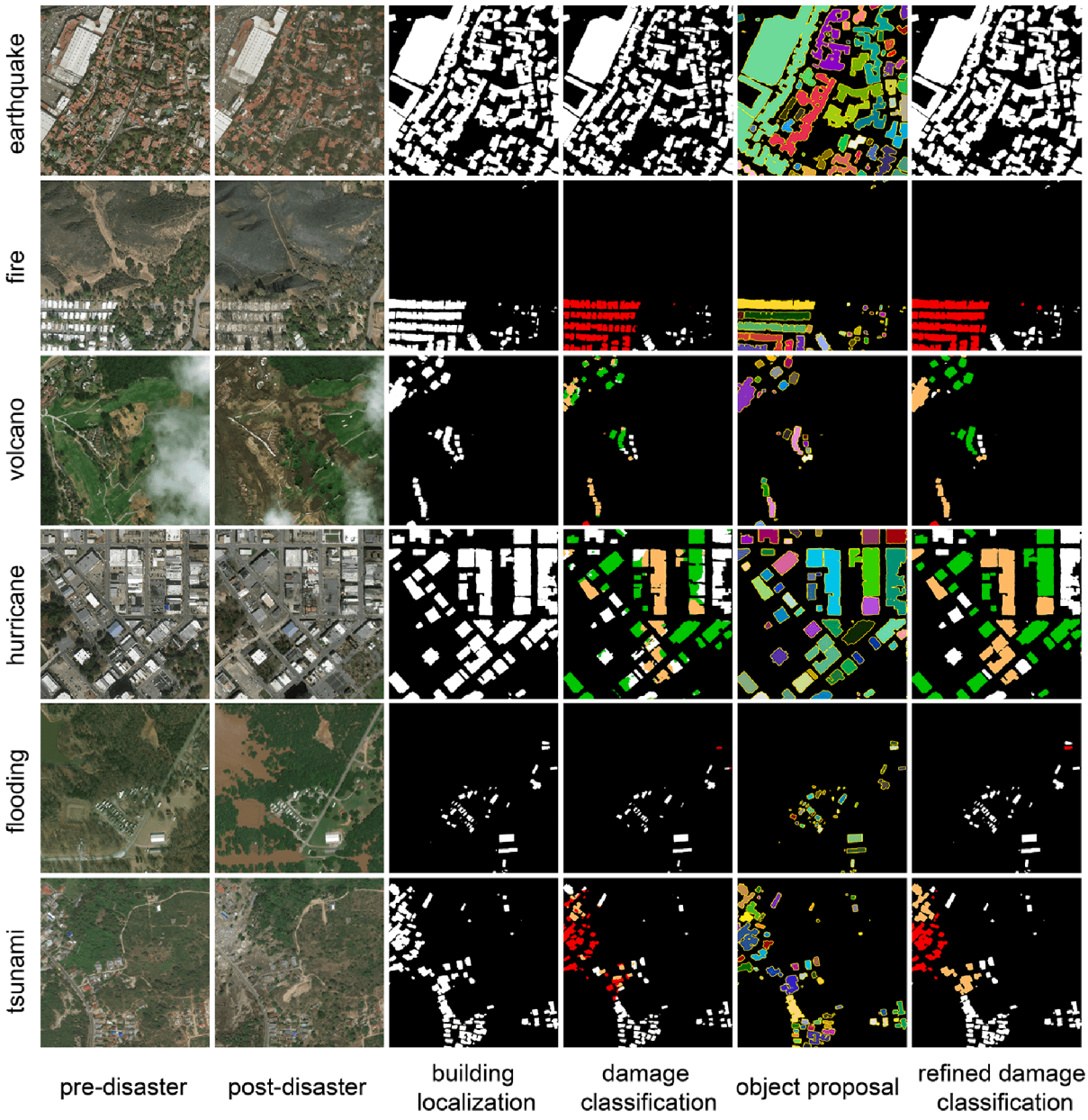

Fig. 8. Intermediate results of the object-based post-processing in ChangeOS.

environment in the city of Beirut.

4.5.3. *Limitations in real-world applications*

In disaster response scenarios, there are low-quality images during disasters due to bad weather conditions. For example, cloud occlusion will make many buildings invisible in the optical satellite images (Zhang et al., 2021). As shown in Fig. 11(a) and (b), buildings under the thick cloud are impossible to be recognized in optical images, which is a limitation of ChangeOS.

5. Conclusion

In the context of complex and diverse natural and man-made disasters, building damage assessment using bitemporal HSR remote sensing imagery is a meaningful but challenging task for the humanitarian assistance and disaster response. The current key problem lies in how to learn a semantically consistent strong feature representation for the building damage assessment. The conventional OBIA framework can guarantee semantic consistency but with weak feature representation, while the Siamese FCN framework has strong feature representation but is semantically inconsistent. In this paper, we have proposed a deep object-based semantic change detection framework, called ChangeOS, to seamlessly integrate OBIA and deep learning to overcome their respective limitations. ChangeOS innovatively integrates building localization and damage classification into a unified end-to-end deep OBIA framework. To make the object segmentation in the framework differentiable, a deep object localization network is adopted to generate accurate building objects in place of the superpixel segmentation commonly used in the conventional OBIA framework. This can also provide deep object features to supply an object prior to the deep damage classification network for more semantically consistent feature representation. ChangeOS also adopts object-based post-processing to further guarantee the semantic consistency for each object. The comprehensive experimental results obtained on a global scale building damage assessment dataset and two local scale building damage assessment datasets of the Beirut port explosion event and the Bata

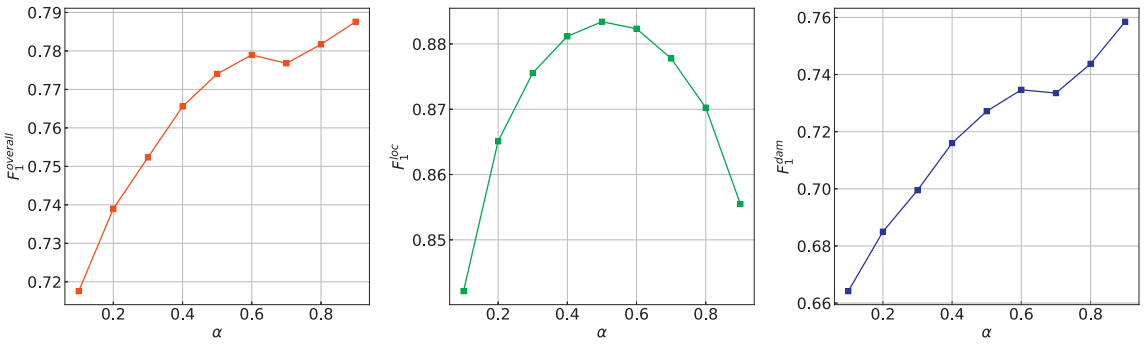

Fig. 9. The impact of the hyperparameter α of Tversky loss on the performance.

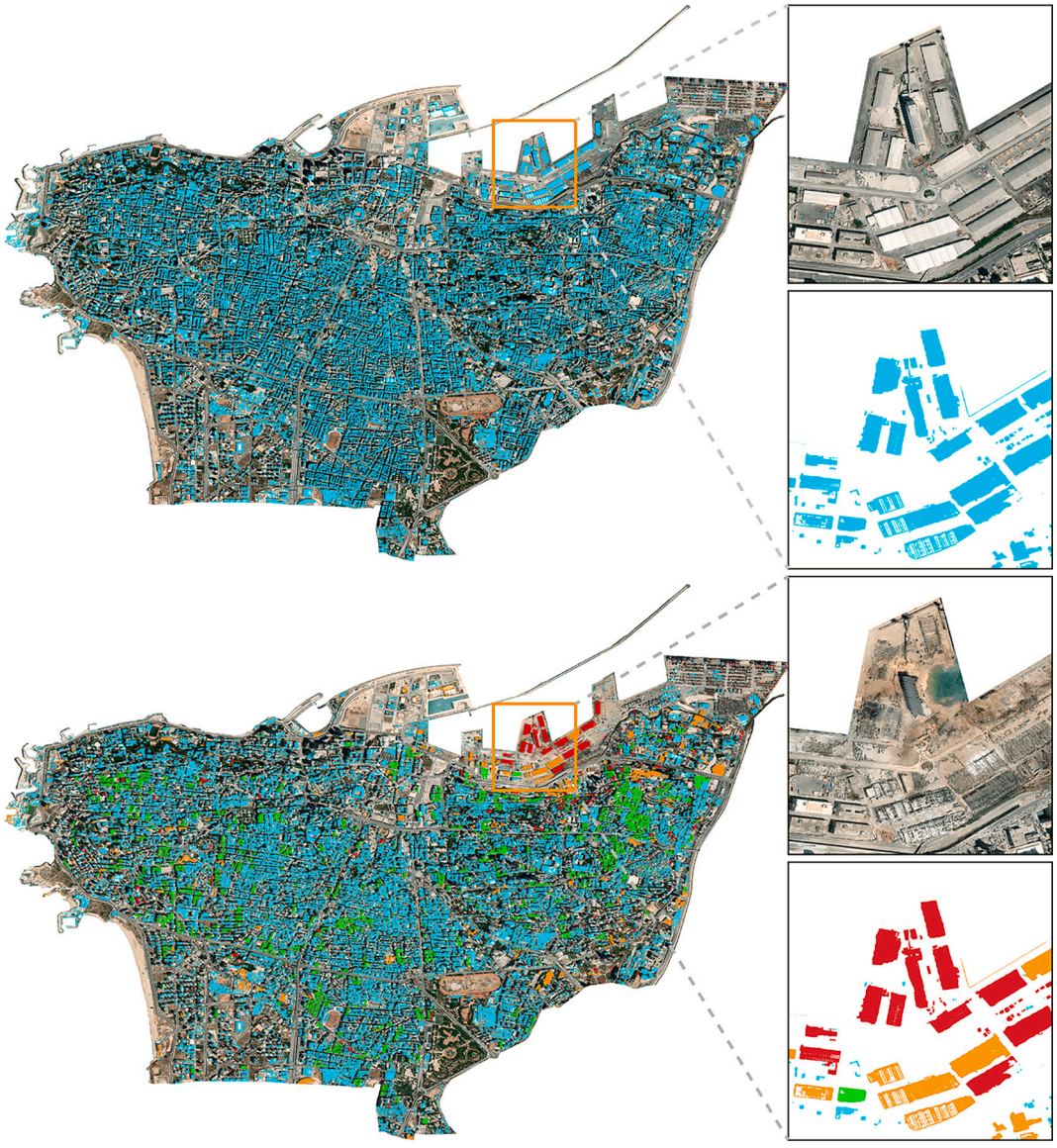

Fig. 10. Visualization of the building damage assessment for the Beirut port explosion event. The minimum bounding rectangle of each image has a size of 11,880 × 16,744. (a) Building localization. (b) Building damage assessment. (c) Sub-region of the pre-disaster image. (d) Sub-region of the post-disaster image. (e) Sub-region of the building localization. (f) Sub-region of the building damage assessment. To improve the visibility, we changed the color of buildings belonging to the non-damage class from white to blue. Legends: Non-damage, Minor damage, Major damage, Destroyed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Building damage assessment performance comparison between UNet+ResNet, Siamese-UNet, and the proposed ChangeOS for the Beirut port explosion event dataset. "e2e" indicates whether the method can perform end-to-end training and inference, and "object-based" indicates whether the method follows the OBIA framework.

Method	e2e	object-based	$F_1^{overall}$ (%)	F_1^{loc} (%)	F_1^{dam} (%)	Damage F_1 per class (%)				Inference time (s)	
						No Dmg.	Minor Dmg.	Major Dmg.	Destroyed	CPU	GPU
UNet + ResNet		✓	12.141	37.413	1.311	14.198	–	0	0.451	1978	–
Siamese-UNet	✓		50.872	56.645	48.398	29.978	–	85.991	58.831	–	927
ChangeOS	✓	✓	53.551	64.498	48.860	27.436	–	89.355	72.676	623	52

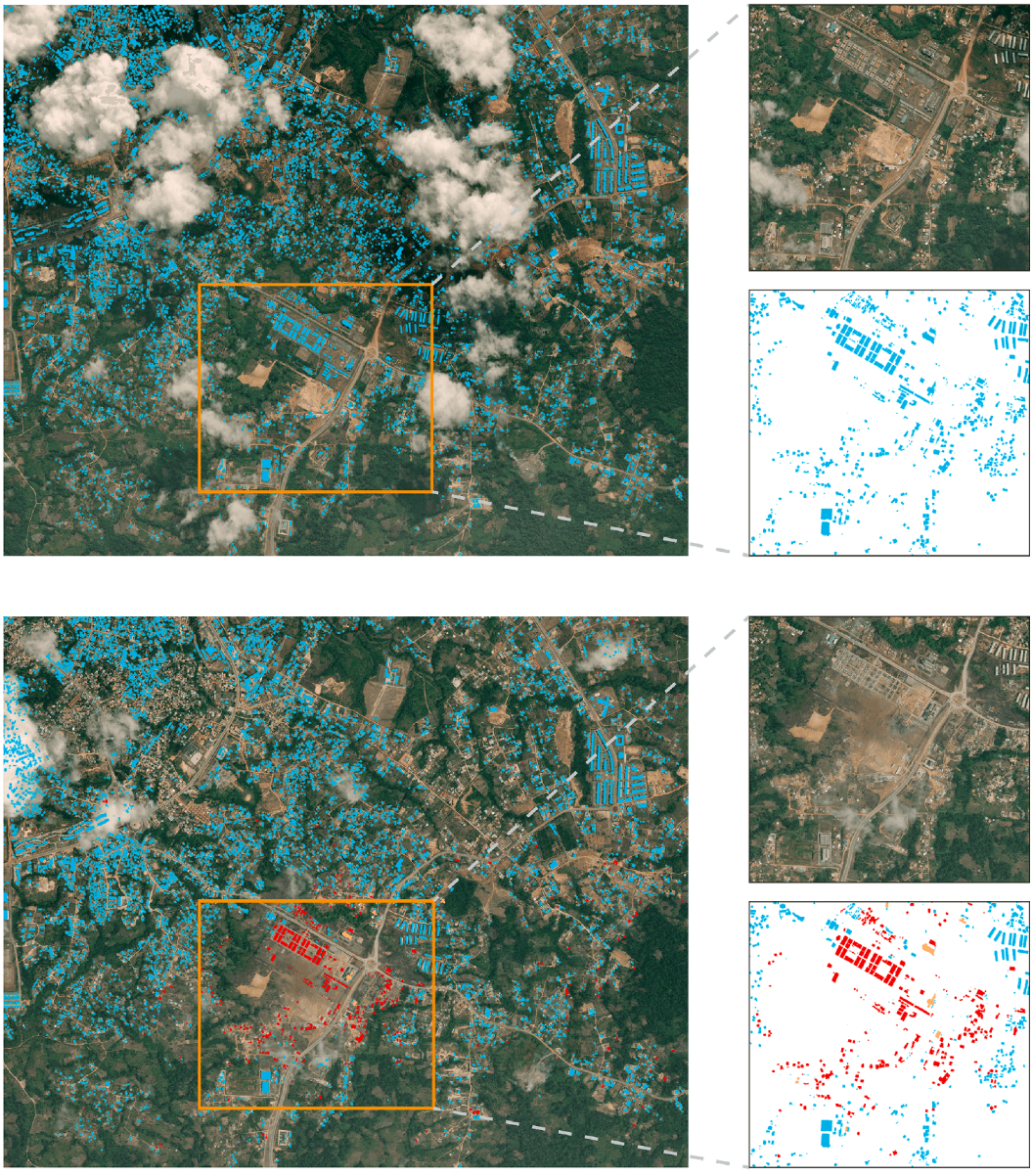

Fig. 11. Visualization of the building damage assessment for Bata military barracks explosion event. Each image has a size of 8085 × 10,033. (a) Building localization. (b) Building damage assessment. (c) Sub-region of the pre-disaster image. (d) Sub-region of the post-disaster image. (e) Sub-region of the building localization. (f) Sub-region of the building damage assessment. For visibility, we changed the color of the buildings belonging to the non-damage class from white to blue. Legends: Non-damage, Minor damage, Major damage, Destroyed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
Building damage assessment performance comparison between UNet+ResNet, Siamese-UNet and the proposed ChangeOS for Bata military barracks explosion event dataset. "e2e" indicates whether the method can perform end-to-end training and inference, and "object-based" indicates whether the method follows the OBIA framework.

Method	e2e	object-based	$F_1^{overall}$ (%)	F_1^{loc} (%)	F_1^{dam} (%)	Damage F_1 per class (%)				Inference time (s)	
						No Dmg.	Minor Dmg.	Major Dmg.	Destroyed	CPU	GPU
UNet + ResNet		✓	55.792	69.972	49.716	50.097	–	–	49.342	970	–
Siamese-UNet	✓		94.581	96.373	93.814	94.597	–	–	93.045	–	342
ChangeOS	✓	✓	97.528	96.986	97.761	96.788	–	–	98.755	247	23

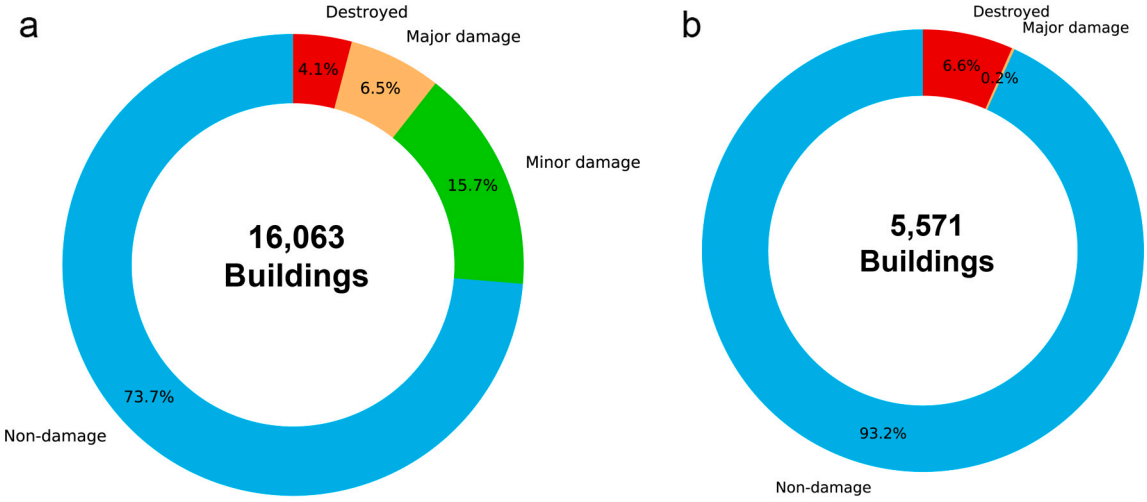

Fig. 12. Statistics of the building states after these two man-made disaster events.

military barracks explosion event show that ChangeOS has excellent performance and has a strong generalization ability. From the ablation study for ChangeOS, we found that the introduction of end-to-end learning is the most important improvement for the building damage assessment. The object prior can also significantly boost the accuracy. We also surprisingly found that the object prior can be obtained from the deep object features, and not just object-based post-processing, both of which can achieve similar accuracy improvement. We believe that ChangeOS could become a strong baseline for building damage assessment, and it represents a robust tool to further promote future research in humanitarian assistance and disaster response.

In the future, we will further study robust building damage assessment in more complex imaging conditions (e.g., thick cloud occlusion) by multi-modal (Zheng et al., 2021) and multi-temporal remote sensing images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the xView2 team for sharing xView2 dataset and Maxar Technologies for sharing satellite images of two local study sites in this study. This work was supported in part by the National Key Research and Development Program of China under grant no. 2017YFB0504202, in part by the National Natural Science Foundation of China under grant nos. 41771385 and 41801267, and in part by the China Postdoctoral Science Foundation under grant no. 2017M622522.

References

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65 (1), 2–16.
Brunner, D., Lemoine, G., Bruzzone, L., 2010. Earthquake damage assessment of buildings using vhr optical and sar imagery. IEEE Trans. Geosci. Remote Sens. 48 (5), 2403–2420.
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.
Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54 (12), 7405–7415.
Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. ISPRS J. Photogramm. Remote Sens. 84, 85–99.
Durnov, V., 2020. xview2 First Place Solution. https://github.com/DIUx-xView/xView2_first_place.
Ge, P., Gokon, H., Meguro, K., 2020. A review on synthetic aperture radar-based building damage assessment in disasters. Remote Sens. Environ. 240, 111693.
Grünthal, G., 1998. European Macroseismic Scale 1998. European Seismological Commission (ESC). Tech. Rep.
Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M., 2019a. Creating xbd: a dataset for assessing building damage from satellite imagery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 10–17.
Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., Gaston, M., 2019b. xbd: A Dataset for Assessing Building Damage From Satellite Imagery. arXiv:1911.09296 (arXiv preprint).
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens. Environ. 214, 73–86.
Kelman, I., 2003. Physical Flood Vulnerability of Residential Properties in Coastal, Eastern England. University of Cambridge. Ph.D. Thesis.
Koshimura, S., Moya, L., Mas, E., Bai, Y., 2020. Tsunami damage detection with remote sensing: a review. Geosciences 5 (1), 177.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, pp. 1097–1105.

Lee, J., Xu, J.Z., Sohn, K., Lu, W., Berthelot, D., Gur, I., Khaitan, P., Koupparis, K., Kowatsch, B., et al., 2020. Assessing Post-Disaster Damage From Satellite Imagery Using Semi-Supervised Learning Techniques. arXiv:2011.14004 (arXiv preprint).

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125.

Liu, T., Yang, L., Lunga, D., 2021. Change detection using deep learning approach with object-based image analysis. Remote Sens. Environ. 256, 112308.

Liu, Y., Chen, D., Ma, A., Zhong, Y., Fang, F., Xu, K., 2020. Multiscale u-shaped cnn building instance extraction framework with edge constraint for high-spatial-resolution remote sensing imagery. IEEE Trans. Geosci. Remote Sens.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). IEEE, pp. 565–571.

Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1520–1528.

Plank, S., 2014. Rapid damage assessment by means of multi-temporal sar-a comprehensive review and outlook to sentinel-1. Remote Sens. 6 (6), 4870–4906.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 (arXiv preprint).

Tong, X., Hong, Z., Liu, S., Zhang, X., Xie, H., Li, Z., Yang, S., Wang, W., Bao, F., 2012. Building-damage detection using pre-and post-seismic high-resolution satellite stereo imagery: a case study of the may 2008 Wenchuan earthquake. ISPRS J. Photogramm. Remote Sens. 68, 13–27.

Tversky, A., 1977. Features of similarity. Psychol. Rev. 84 (4), 327.

Valentijn, T., Margutti, J., van den Homberg, M., Laaksonen, J., 2020. Multi-hazard and spatial transferability of a cnn for automated building damage assessment. Remote Sens. 12 (17), 2839.

Vickery, P.J., Skerlj, P.F., Lin, J., Twisdale Jr., L.A., Young, M.A., Lavelle, F.M., 2006. Hazus-mh hurricane model methodology. ii: damage and loss estimation. Nat. Hazards Rev. 7 (2), 94–103.

Wu, K., Otoo, E., Shoshani, A., 2005. Optimizing connected component labeling algorithms. In: Medical Imaging 2005: Image Processing. vol. 5747. International Society for Optics and Photonics, pp. 1965–1976.

Yamazaki, F., Matsuoka, M., 2007. Remote sensing technologies in post-disaster damage assessment. J. Earthq. Tsunami 1 (03), 193–210.

Yusuf, Y., Matsuoka, M., Yamazaki, F., 2001. Damage assessment after 2001 gujarat earthquake using landsat-7 satellite images. J. Indian Soc. Remote Sens. 29 (1–2), 17–22.

Zhang, C., Harrison, P.A., Pan, X., Li, H., Sargent, I., Atkinson, P.M., 2020. Scale sequence joint deep learning (ss-jdl) for land use and land cover classification. Remote Sens. Environ. 237, 111593.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (ocnn) for urban land use classification. Remote Sens. Environ. 216, 57–70.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2019. Joint deep learning for land cover and land use classification. Remote Sens. Environ. 221, 173–187.

Zhang, Q., Yuan, Q., Li, Z., Sun, F., Zhang, L., 2021. Combined deep prior with low-rank tensor svd for thick cloud removal in multitemporal images. ISPRS J. Photogramm. Remote Sens. 177, 161–173.

Zheng, Z., Ma, A., Zhang, L., Zhong, Y., 2021. Deep multisensor learning for missing-modality all-weather mapping. ISPRS J. Photogramm. Remote Sens. 174, 254–264.

Zheng, Z., Zhong, Y., Ma, A., Han, X., Zhao, J., Liu, Y., Zhang, L., 2020a. Hynet: hyper-scale object detection network framework for multiple spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 166, 1–14.

Zheng, Z., Zhong, Y., Ma, A., Zhang, L., 2020b. Fpga: fast patch-free global learning framework for fully end-to-end hyperspectral image classification. IEEE Trans. Geosci. Remote Sens.

Zheng, Z., Zhong, Y., Wang, J., Ma, A., 2020c. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4096–4105.

Zhong, Y., Han, X., Zhang, L., 2018. Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 138, 281–294.